SEEDS Standards and Interfaces Process Study Task

Plenary Overview

Ken McDonald

Rich Ullman

NASA Goddard Space Flight Center

SEEDS Community Workshop,
June 17-19, 2002

Agenda

- **Motivation for standards and standards processes**
  - (pick up on Silvia's talk)
  - What is enabled by standards?
  - Sources of requirements for SEEDS
- **Near Term Standards Status**
  - Overview of study task
  - Summary of Results
- **Long Term Standards & Interfaces Processes Status**
  - Overview of study task
  - Summary of Results
- **Breakout plans and goals**

Premise

- "An important premise underlying [SEEDS] is that its various parts should have considerable freedom in the ways in which they implement their functions and capabilities. Implementation will not be centrally developed, nor will the pieces developed be centrally managed. However, every part of [the ESE network] should be configured in such a way that data and information can be readily transferred to any other. *This will be achieved primarily through the adoption of common interface standards and practices* [1]."
- Standards are what make increased flexibility for SEEDS participants possible.

Why Standards?

- To increase the use of the ESE data products for Earth science research and applications.
- To facilitate interoperability between different components of the ESE network of data systems.
- To make it easy for data and service providers to join the ESE network of data systems.
- To reduce costs for the ESE network of data systems as a whole.

Why Are SEEDS Standards Processes Needed?

- SEEDS is expected to consist of a mix of loosely coupled diverse, distributed components derived from the contributions of numerous individual investigators, data providers, and institutions.
- The SEEDS components, while loosely coupled, are expected to conform to a minimal set of Core interfaces and standards
- The SEEDS community is expected to participate in the processes that will determine what the SEEDS Core interfaces and standards are

When a new interface is needed that is not covered by a SEEDS standard, look first for an international, national, federal, or defacto standard to adopt, profile, or extend. If none available that meet mission schedules and requirements, then develop custom NASA (SEEDS) interface standard

Drivers for SEEDS Core Interfaces and Standards?

- Science Data Interuse
- Data Access Interoperability
- Applications Support
- HQ mandate
- Federal Mandate
- Agreements with other US and foreign agencies

How these core interfaces and standards are identified will be worked out with community input

Near Term Standards Section

- Do you want to put an updated Near Term Standards Quad chart here?

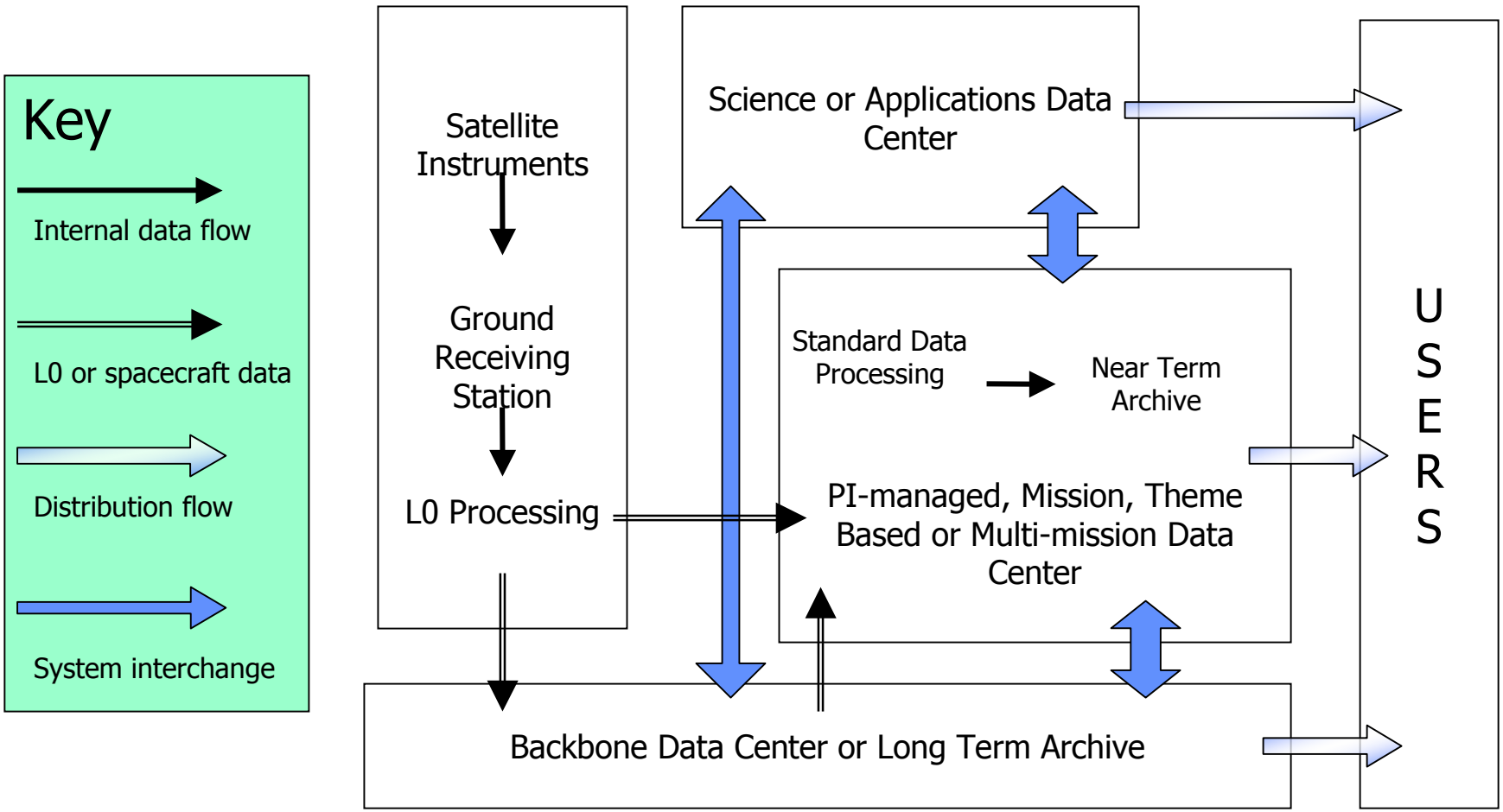
Purpose of the Near-Term study

- **Investigate standards for the ESE network of data systems and services for near-term missions:**
  - Data distribution packaging standards
  - Data interchange packaging standards
  - Metadata and documentation standards
  - Data interface standards

Why Near-Term Missions Study?

- Near-term missions are missions already in formulation; however, for SEEDs guidance to be useful, it needs to be specific and soon.

Simplified Data Flow Diagram



NTMS Study Methodology

- Survey near-term missions and heritage missions
- Conduct user interview and survey
- Survey and critique standards currently in use in heritage missions or under consideration by near-term missions
- Gather lessons learned from heritage NASA and NOAA missions
- Perform standards analysis
- Evaluate technology solutions for interoperability

Near-Term Missions

Mission Name	Phase	ANTICIPATED Launch Date
Landsat Data Continuity Mission (LDCM)	Formulation	2005
NPOESS Preparatory Project (NPP)	Formulation	2005
Ocean Surface Topography Measurement (OSTM)	Formulation	2005
Ocean Vector Winds	Formulation	2006
Global Precipitation Measurement (GPM)	Formulation	2007
Solar Irradiance	Formulation	2007
Carbon Cycle Initiative (CCI)	Pre- Formulation	2008-2012
Total Column Ozone	Pre- Formulation	N/A

Summary NTMS heritage standards

- **Data packaging standards**
  - netCDF, GeoTIFF, Fast Format, Binary, BUFR, HDF, HDF-EOS
- **Metadata standards**
  - ECS data model, FGDC content standard, GCMD DIF
- **Documentation standards**
  - EOSDIS Guide

Standards Evaluation Criteria

- **The study team evaluated and surveyed UWGs about identified standards according to many "-ility" criteria. These criteria are indicative rather than exhaustive.**
- **The intention is not to identify one all purpose standard, but rather to identify appropriate use.**
- **Two kinds of interfaces that can benefit from ESE standards:**
  - For distribution, standards must be acceptable to target community,
  - For interchange among systems, standard must have semantic completeness, descriptive power and portability.

User Interview/Survey

- 20 data producers/users from the NASA Science Data Processing Workshop 2002
- 25 DAAC User Working Group members and other users
- All of the users/producers answered questions related to data format standards. Only one-quarter of the users/producers answered questions related to metadata standards.
- Many interviewed/surveyed are not familiar with multiple data formats, with only one-half of the respondents familiar with more than two data standards. In many cases they gave the standard(s) they were most familiar with the highest rating.

Interview/Survey Results

- **Data packaging standards:**
  - HDF and netCDF were rated highest for Portability and Suitability.
  - "Binary" format (i.e. product specific) received the highest ratings for Availability, Interoperability, and Evolvability.
  - 60% of respondents had success with HDF; 33% recommend HDF as a future standard for NASA. 25% say HDF was an impediment to their work.

Interview/Survey Results (cont.)

- **Metadata standards:**
  - The ECS data model and the FGDC content standard received comparable ratings for most criteria.
  - For future metadata standards, some recommend ISO 19115 to replace the current FGDC, and FGDC extensions for remote sensing based on the ECS data model.
  - Others recommend XML standard descriptions for metadata and refining the ECS data model.

Key NTMS Findings(1)

- ESE Data Systems must provide many more options for data packaging, even in the near term.
- In the near term, the chief mode of delivering data remains the transfer of discrete files. The use of Web Services is still only emerging.
- In near term, content data standards alone may not suffice for transferring complex data between different user communities without information loss or corruption.
- Formats will become even less important as application-to-application interfaces evolve making formats invisible to the producer and consumer.

Key NTMS Findings(2)

- **Requirements for interchange among ESE components are different from requirements for distribution to end users.**
- **Missions must plan for evolution of end user requirements for packaging of mission science data over the life time of the mission.**
  - Data formats
  - Data distribution system interfaces
  - Associated metadata

Long Term section

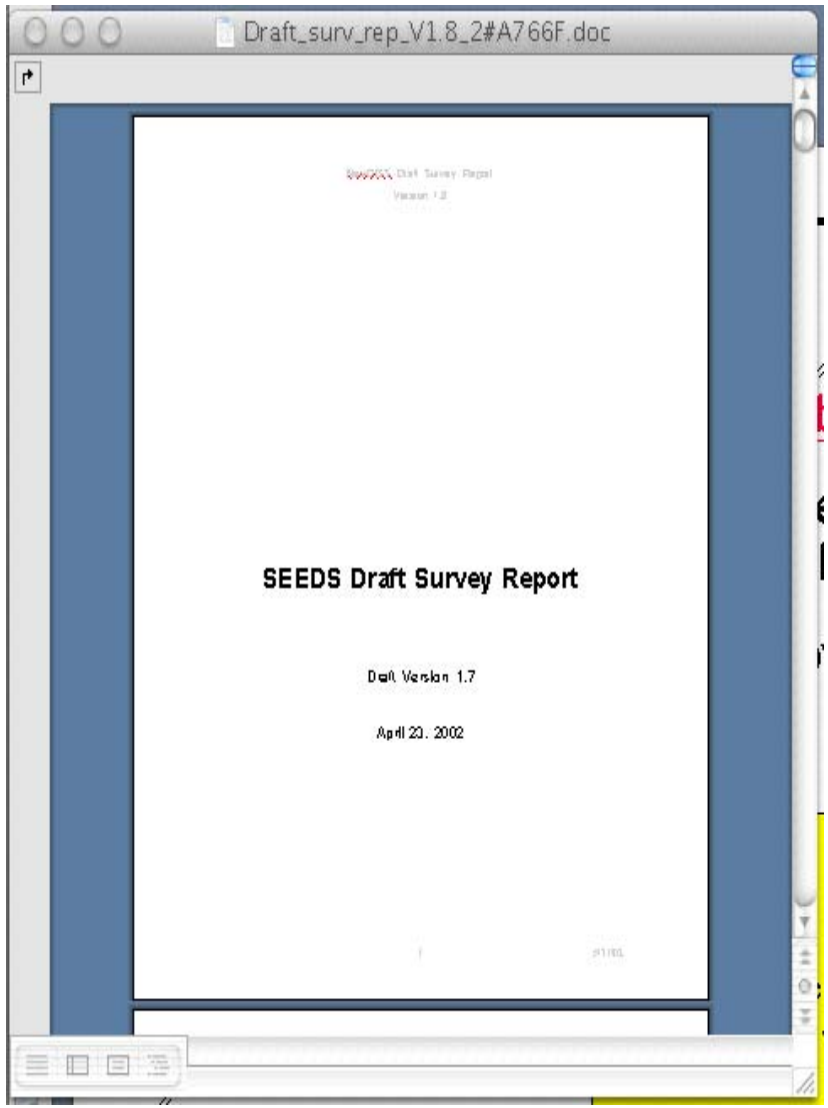
- Do you want to put in the Long Term Standards and Interfaces Processes team quad chart here?

Purpose of the Standards Processes Study

- **Recommend processes for developing and approving standards for the ESE network of data systems and services throughout the SEEDS era.**
- **Types of standards to be addressed by these processes:**
  - Distributed information search protocols (e.g., EOSDIS V0, Z39.50)
  - Data interface standards (e.g, OGC Web Map Services, DODS)
  - Data distribution packaging standards (e.g., HDF-EOS)
  - Data interchange packaging standards (e.g., HDF-EOS)
  - Metadata and documentation standards (e.g., GCMD DIF, ESML)
  - Service documentation standards (e.g., GCMD SERF)
  - Service communication protocols (e.g., UDDI, WSDL, SOAP)

Draft Survey Reports

- **Get reports here: http://lennier.gsfc.nasa.gov/seeds**
  - The reports' content is reflected in the notional process ideas presented at this workshop.



Survey Report - Section 1 Introduction

- **The report strives to answer 2 questions:**
  - How were standards adopted within ESE and in similar programs in other agencies in the past and how successful was the adoption in terms of actual implementation experience?
  - What are some of the formal standards bodies that produce standards relevant to ESE, what are their internal processes, and how can ESE benefit from participating in these standards bodies, particularly in light of the experiences captured in the first part?
- **These questions are asked because we know that SEEDS systems will use standards in one of at least 4 ways:**
  - Adopt : Adopt a standard as-is and use it. E.g. OGC Web Map Server
  - Profile : Adopt a standard but constrain its use in some way. E.g. FGDC metadata content has many optional fields. SEEDS could decide which ones to use/not use
  - Extend : Adopt a standard but extend its use in some way. E.g. SMTP (Simple Mail Transport Protocol) allows for extensions in the mail headers.
  - Develop : Develop a standard for use within SEEDS. E.g. SEEDS may want to develop a standard mechanism for distributing new versions of controlled vocabularies for metadata entries.
- **The document provides an overview of standards types and factors that should be considered in the process of adoption or development.**

Survey Report - Section 2
NASA-ESE experiences with Standards and Interfaces

- **Key factors observed for widespread adoption of a standard or interface within the NASA EOSDIS:**
  - Community involvement during the development process
  - A small group of people involved during development process
  - Strong project management staff needed to lead the technical discussions, the implementation, and overall management essential
  - Software tools and components readily available. This may require NASA investment for user/research communities to develop tools and/or provide technical support.
  - Simple interfaces
- **Based on these factors, overall recommendations are made. These are incorporated into the notional process discussions.**

Survey Report - Section 3
Other experiences with standards & interfaces

- **Offers some points of comparison with NASA's own experiences in terms of creating and deploying information standards and interfaces.**
- **Reviews the experiences of**
  - Canada's GeoConnections program
  - US National Oceanographic and Atmospheric Administration (NOAA)
  - Global Grid Computing initiative
  - Sun's Java Community Process.

Survey Report - Section 4
Experiences of Standards Organizations

- **This section looks at some of the major standards bodies whose output is used within ESE and attempts to describe the kinds of standards they produce as well as the processes they follow internally.**
- **The organizations described are:**
  - ISO TC 211 Geographic Information/Geomatics
  - Open GIS Consortium (OGC)
  - World Wide Web Consortium (W3C)
  - Consultative Committee for Space Data Systems (CCSDS)
  - Federal Geographic Data Committee (FGDC)
  - Internet Engineering Task Force (IETF)

Breakout Plan and Goals

- **Two identical, parallel standards breakouts Tuesday afternoon to enable small group discussion**
- **Discussion of results for Near Term Standards study**
  - Lessons Learned
  - Standards evaluation criteria & Analysis of results
  - User interview/survey results
- **Discussion of results for Long Term Standards and Interfaces Processes**
  - Overview of study findings and identification of different processes
  - Observations and lessons learned
  - Analysis and strawman processes and activities
- **Participants and Their Roles in Processes and Associated Activities**
  - Identification of communities and stakeholders
  - What are roles with respect to various activities
  - What is NASA's role?
  - Discuss SEEDS relationship and interactions with other groups